ARE YOU DATA-READY FOR THE AGE OF AI?

INSIGHTS

- Even though the investment in AI has significantly increased in 2024, an **Infosys survey** revealed that AI outputs failed to excite 74% of respondents, with AI being perceived as **effective merely 25% of the time.**

- Businesses need to be data-ready to make the most of the AI revolution.

- We propose a blueprint for getting enterprise data ready for AI.

- This blueprint focuses on intelligent systems and automation solutions for data management and decision-making.

In the recent past, boardroom discussions focused on whether AI's elusive promises could be realized. But a lot has changed since then. From end-to-end AI-powered automated factory floors and drone-based product delivery to bot-supported surgeries at hospitals, AI is revolutionizing business operations, unlocking new levels of productivity and growth.

It is no longer a question of whether we can achieve the full potential of AI; rather 'how do we best leverage the power of AI to drive business value at scale?'

However, despite the best attempts, many businesses still need to catch up in the pursuit of AI excellence. Even though there's **a 67% increase in generative AI investments** alone from 2023 to 2024, 74% were not impressed by the AI outputs, and only 25% of the enterprises are satisfied with these production-ready initiatives.

What's causing this pilot purgatory? As AI evolves beyond individual applications to target enterprise-scale transformations,

there is a need for a significant shift in how we look at the relationship between AI and data. AI is only as good as the data used to train it. And unlike other models, AI is non-transparent, increasing the need for data reliability.

From model-driven to data-driven: It's time for a new approach.

We are living in an era of profound change in AI functionalities. Formerly, we relied solely on models built strictly on traditional algorithms. This resulted in erroneous actions if the AI encountered a scenario outside the parameters the model was built to handle. However, the rapid progress in this field has given rise to a newer generation of algorithms that are trained on large data sets to learn different patterns, relationships, and behaviors. Therefore, AI currently has become more dependent on large-scale data and can think outside the strict parameters of a traditional model.

The success of such large-scale data- driven AI models depends on the quality, quantity, and variety of data that is used to train it and/or provided to it for inferencing. The quality of input data directly affects the output result that one can expect, and in the case of generative AI, catching these drifts in tests and correcting for them, is all the more difficult. AI used to be purpose-built for particular use cases, but with the advent of generative AI, all of the enterprise data could be used to train large language models or provided to them for inferencing, including user-generated content, meeting notes, audio and video, etc. Making this enterprise data available in a safe, secure, and privacy-preserving manner is a task that is easier said than done, given how siloed and fragmented traditional data solutions are today. Certain roadblocks need to be considered.

Challenges and considerations of "large-scale data-driven AI"

Enterprises have been largely used to managing analytical and historical data from a business intelligence perspective. With the advent of large-scale data-driven AI, the universe of data expands significantly to user-defined data, audio, video, transactional, and 3rd party data. Traditional data platforms have to expand their capability to store, manage, and process this universe of data in a security and privacy-preserving way to produce intelligence that can be trusted and clearly quantifies the risk of its usage.

> **"Enterprises need to be able to store, manage, and process the expanding universe of data including user-generated content along with access rules and privileges that come with that."**

Data preparation is one of the biggest barriers to unleashing the might of the current generation of AI. A primary challenge in data preparation? Collecting and labeling data from various touchpoints. Today, every business is expanding both its geographic reach and customer base. As a result, they are dealing with a myriad of data points covering employee, operational, transactional, customer information, and more. In addition to cleaning this data, which itself is an onerous task, we need to make sure the data is tagged with the appropriate usage, confidentiality, and privacy considerations to be able to comply with regulatory and enterprise rules of fairness and trust when training or leveraging your AI applications. Doing this on a case-by-case basis is an extremely expensive and resource-intensive task.

> **"Data preparation is a crucial barrier to success for many businesses. Doing this on a case-by-case basis is extremely expensive, given the duplication, and unsuitable for a ubiquitous AI environment."**

Organizing the data into schemas and labeling it through metadata eases the consumption of this data to train data-driven AI models. While the organization and connectedness of data can be through semantic models, knowledge graphs, or data products, the consumption of data is an important consideration that accelerates the training of AI and the subsequent ability to scale AI initiatives in the enterprise. Lack of consideration of the storage, organization, connectedness, and consumption of data becomes a major challenge in the adoption and scaling of large-scale data-driven AI in an enterprise.

> **"Lack of focus on the consumption and connectedness of the data is a major barrier."**

Data management is another area that impedes many organizations from becoming data-ready. This includes everything from storage, security, privacy, and metadata management through information lifecycle management, which includes storage and destruction of data. This is the lever that allows us to maintain the integrity of the data corpus that is used for training the AI or provided to AI for inferencing. It averts the risk of cybersecurity breaches and makes sure that the AI is trained with the right type of data that does not violate confidentiality and privacy boundaries.

Focusing on good data governance practices by the enterprise, leveraging AI/ML itself to do it at scale, is the answer to this, and it is able to sustain the value that you can get from your data and AI initiatives. Consequently, breaking away from siloed, inconsistent, and erroneous data.

Data governance is critical for other reasons as well: Handling a vast amount of data, encompassing personal, private, and sensitive organizational information, comes with a huge responsibility. Businesses are required to comply with various data protection laws and regulations, such as GDPR, CCPA, and HIPAA, as well as maintain trust with all stakeholders. Good data governance practices can help businesses comply with these regulatory mandates.

> **"Good data and AI governance is key to getting sustained value from your data and AI initiatives."**

In addition to this, there are a few areas that hold organizations back from being data-ready for AI:

- **Unclear business strategy for AI**
  Many businesses try incorporating AI and generative AI solutions into their workflow without a proper strategy and focus. This results in improper planning for the data needed to support the objectives, resulting in the unrealized potential of these solutions.

- **Lack of trust in AI control**
  Despite its popularity, there is still some skepticism about AI. Organizations need to sensitize their employees with proper workshops to improve their reception of any operational change and ensure their participation.

- **High TCO and uncertain ROI**
  Many organizations also face financial challenges. Establishing proper data governance in-house requires a significant financial investment. Given the huge investment businesses have had to make, it can often result in low and uncertain ROI if sight to value is not maintained.

Stepping into a new era

Tackling these challenges, businesses are at different stages in their AI journey. A recent IDC report states that almost 70% of the current AI market value is dedicated to making businesses data-ready for AI. Even our own **market research** supports this data.

We believe a **five-dimensional framework** focusing on people, processes, platforms, ecosystems, and data can help enterprises **become data-ready.** Such a holistic outlook helps organizations become responsible by design and establish the right control functions for better data governance.

Businesses can set up an end-to-end data estate that ensures all data assets are always accessible and of high quality. Once the data estate is set up, they can organize and fingerprint the data so that it is properly cataloged with a common understanding for AI initiatives. For e.g., user-generated content such as email, notes, video, and audio need to be fingerprinted for confidentiality, privacy, and compliance so that they can be used appropriately when training both regular and generative AI systems. It gives us the ability to provide access control for all organizational data, synthetic data versioning, and protection against various security concerns.

As we connect the data and start to look at vertically integrated workflows across the silos of the company, AI that learns from this data can become more effective and productive across this workflow.

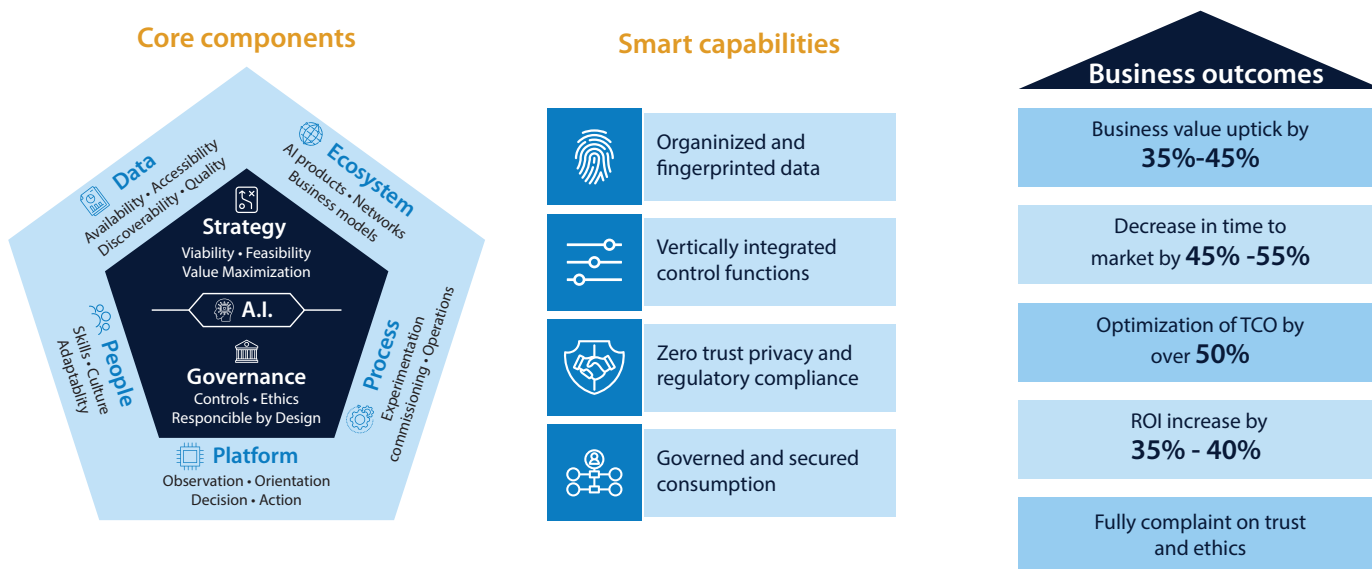**This will be the master data & AI blueprint across tech stacks**

Figure 01: Blueprint for getting enterprise data ready for AI

This five-dimensional blueprint underscores the end-to-end AI narrative for many market leaders. As a data-ready entity, businesses employ smart AI and generative AI capabilities across their workflow to drive significant business value. In using enterprise unstructured data, data privileges, and insights from the wisdom of crowds to avoid random noise, are key aspects that are leveraged in this approach.

The focus on governed data and responsible AI increases the security, privacy, and trust in the system, making the results from the use of AI ethical, trustworthy, and explainable. The ability to comply with regulatory requirements like GDPR, especially for user-generated content, and operating under a zero-trust policy make these solutions much more reliable. To become responsible by design, enterprises can follow the AI3S framework (scan, shield and steer) to easily identify and improve the overall risk posture, legal obligations, vulnerabilities, and threats arising due to AI adoption using a single source of truth for tracking the risk and compliance status of all AI projects.

"Data-ready for AI primarily leverages data fingerprinting, connected data across silos of the enterprise, the wisdom of crowds in using unstructured data, auto-checks of data imbalance and bias, and the scan-shield-steer framework for AI risk management."
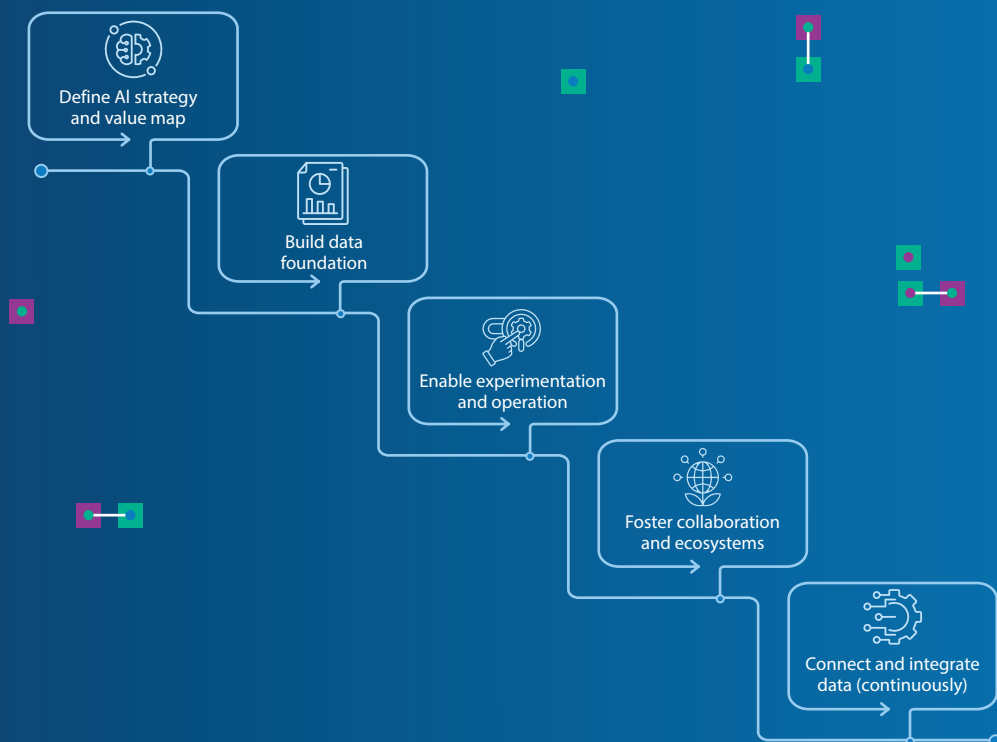
Preparing with data-readiness

Getting the enterprise data ready for AI can be done in stages driven by the company's overall AI strategy. We will need to look at the objectives of the company and how AI can help. Based on this, a value map needs to be created to look at the business value delivered by each of the AI initiatives, whether top-down or bottom-up. This can guide the prioritization of the areas of the data foundation and platform that would deliver the maximum business value. This will help you progressively build out the data platform that will allow for experimentation and smart operations at its core. People will need to be trained to understand the AI-first approach and the use of such a platform for experimentation and execution. Finally, extending the platform to enable the sharing of intelligence across organizational boundaries helps with the creation of ecosystems, including your partners, to create new offerings and address new business opportunities.

Figure 02: 5 steps to become data-ready

- Define AI strategy and value map
- Build data foundation
- Enable experimentation and operation
- Foster collaboration and ecosystems
- Connect and integrate data (continuously)

Case study

From Soft Beverage retailer to F&B experience center

The journey of a company from a beverage retailer to a food and beverage experience center underscores getting enterprise data ready for the AI journey. Below is the roadmap of the customer's journey through their long-standing partnership with Infosys:

A TIMELINE

MILESTONE NO.1
Built a data platform using analytical and historical data

MILESTONE NO.2
Launched a customer app to create user-generated content

Using all this data, our client was able to understand customer preferences and choices at the local or micro-district level

MILESTONE NO.3
Our client decided to pivot to an AI-first growth strategy

MILESTONE NO.4
Our client successfully launched a mobile wallet and created more customer data points

MILESTONE NO.5
Our client introduced significant automation and AI-capabilities to its workflow

MILESTONE NO.6
The AI-first strategy allowed them to invest in an expanding ecosystem of services

Figure 03: Our client's success journey powered by data and AI

Data is the foundation for the AI construct

While AI is increasingly becoming an imperative in achieving business excellence, it is time to think beyond and focus on 'how well you use the technology.' However, to be data-ready and drive end-to-end AI implementation, having access to the right partner, outcome-driven strategy, and governance are crucial. We understand that being data-ready is a key challenge for many enterprises vying for an AI-first strategy.

Infosys Topaz is an AI-first set of services, solutions, and platforms that use generative AI technologies. As an extension of your existing resources, we can help you become data-ready and prepare for the ongoing AI revolution.

Explore how **Infosys Topaz** can help you become data-ready for AI.

Infosys Topaz is an AI-first set of services, solutions, and platforms using generative AI technologies. It amplifies the potential of humans, enterprises, and communities to create value. With 12,000+ AI assets, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale, and connected ecosystems. Connect with us at **infosystopaz@infosys.com.**

For more information, contact askus@infosys.com

Infosys®
Navigate your next

Infosys.com | NYSE: INFY

Stay Connected